

PREDIKSI MASA STUDI MAHASISWA DENGAN VOTING FEATURE INTERVAL 5 PADA APLIKASI KONSULTASI AKADEMIK ONLINE

Andre Widjaya, Lely Hiryanto, Teny Handhayani

Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara

Jl. Let. Jend. S. Parman No. 1, Jakarta 11440 Indonesia

Email: liangz_joy@hotmail.com

Abstrak

Aplikasi prediksi masa studi mahasiswa merupakan aplikasi yang mengimplementasikan metode Voting Feature Interval 5 pada fitur prediksi masa studi mahasiswa. Hasil prediksi masa studi akan digunakan sebagai masukan bagi dosen penasehat akademik dan ketua program studi untuk menyusun rencana studi yang tepat bagi mahasiswa. Studi kasus yang digunakan dalam aplikasi adalah prediksi masa studi mahasiswa Program Studi Teknik Informatika, di salah satu perguruan tinggi swasta di Jakarta, angkatan 2013 dan 2014. Adapun data pembelajaran yang digunakan adalah data nilai mahasiswa, program studi Teknik Informatika di perguruan tinggi swasta tersebut angkatan 2008 sampai 2012. Metode Voting Feature Interval 5 bekerja dengan mengklasifikasikan data nilai mahasiswa berdasarkan nilai vote yang terbentuk dari hasil pembelajaran data latih. Hasil pengujian aplikasi menunjukkan bahwa aplikasi berhasil menghasilkan prediksi masa studi bagi mahasiswa dengan tingkat akurasi yaitu sebesar 73,33%.

Kata kunci: *Voting Feature Interval 5, Prediksi Masa Studi Mahasiswa, Klasifikasi*

Abstract

Period of study prediction application is an application that implement Voting Feature Interval 5 method for its students' period study prediction feature. The result of the prediction will be used by head of department and academic advisors to arrange a proper study plan for the student. Case study that was used in this application is period of study prediction for Computer Science department, Tarumanagara University student, class of 2013 and 2014. Learning data that used was mark of selected course from Faculty of Information Technology, Tarumanagara University student, class of 2008 until 2012. Voting Feature Interval 5 method classify student grade in testing dataset based on voting grade formed by learning dataset. The result of testing is fairly accurate with 73,33% in average

Keywords: *Voting Feature Interval 5, Classification, Period of study prediction*

1. PENDAHULUAN

1. Latar Belakang Pemasalahan

Mahasiswa yang lulus dengan masa studi 7 semester, 8 semester dan terlamat yaitu di atas 8 semester cenderung memiliki nilai yang seragam. Adapun keberhasilan mahasiswa

menyelesaikan masa studinya tidak lepas dari faktor adanya rencana studi yang disusun oleh ketua program studi dan dosen penasehat akademik dari mahasiswa yang bersangkutan. Oleh karena itu dibutuhkanlah suatu bahan masukkan bagi ketua program studi dan dosen penasehat akademik untuk menyusun rencana studi yang tepat bagi mahasiswa yang dikemas dalam web konsultasi akademik online.

Aplikasi yang dibangun adalah aplikasi konsultasi akademik online dengan fitur utama melakukan prediksi masa studi mahasiswa dengan metode Voting Feature Interval 5 ke dalam 3 kelas yaitu kelas dengan masa studi 7 semester, 8 semester dan terlambat (di atas 8 semester) untuk program studi Teknik Informatika di salah satu perguruan tinggi swasta di Jakarta.

Pembatasan ke dalam 3 kelas ditentukan berdasarkan kemungkinan yang ada dalam masa studi, yaitu lulus tepat waktu (8 semester) dan lulus terlambat (di atas 8 semester). Sedangkan pemisahan masa studi 7 semester dengan masa studi tepat waktu didasarkan pada asumsi bahwa untuk lulus dengan masa studi 7 semester dibutuhkan effort yang lebih dibandingkan effort dengan untuk lulus dengan masa studi 8 semester.

Pemilihan metode Voting Feature Interval 5 didasarkan pada kelebihan aplikasi yaitu mampu melakukan klasifikasi dengan cepat dan sederhana karena metode juga mampu mengabaikan missing value sehingga pemrosesan data lebih cepat. Hal ini berbeda dengan algoritma klasifikasi lainnya seperti decision tree atau nearest neighbour yang mengharuskan pencarian nilai dari data yang memiliki nilai kosong atau missing value.

Proses klasifikasi diawali dengan pembelajaran dari data latih yaitu data nilai mahasiswa angkatan 2008 sampai 2012. Hasil dari pembelajaran data latih adalah interval yang terisi lengkap dengan frekuensi range interval dan point interval yang sudah dinormalisasi. Frekuensi tersebut nantinya akan digunakan sebagai nilai vote untuk setiap kelas. Kelas dengan nilai vote terbesar menyatakan bahwa kelas tersebut adalah kelas dari data yang dimasukkan. Sedangkan untuk data uji yang digunakan adalah mahasiswa angkatan 2013 dan 2014.

2. PENELITIAN TERKAIT

2.1 Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining

Penelitian terkait yang sudah pernah dibuat adalah “Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining” oleh Muhammad Syukri Mustafa dan I Wayan Simpen, mahasiswa Teknik Informatika STMIK Dipenegara [6]. Aplikasi yang dihasilkan pada penelitian mereka, mengklasifikasikan attribute meliputi nilai ujian nasional (UN), asal sekolah, daerah, jenis kelamin, pekerjaan dan penghasilan orang tua dan jumlah saudara. Metode yang digunakan pada aplikasi mereka adalah metode *K Nearest Neighbour*, sehingga dapat menghasilkan suatu prediksi berdasarkan kedekatan data pembelajaran yang sudah diklasifikasi dengan data pengujian, apakah mahasiswa tersebut berpeluang untuk menyelesaikan studi tepat waktu atau tidak. Dari hasil pengujian dengan menerapkan algoritma *KNN* dan menggunakan data sampel alumni tahun wisuda 2004 sampai dengan tahun wisuda 2010 untuk percobaan pertama dan data alumni tahun wisuda 2011 untuk percobaan kedua diperoleh tingkat akurasi sebesar 83,36%

2.2 Aplikasi Perencanaan Studi Mahasiswa Menggunakan *Bayesian Belief Networks*

Penelitian terkait selanjutnya yang pernah disusun adalah “Aplikasi Perencanaan Studi Mahasiswa Menggunakan *Bayesian Belief Networks*” pada Fakultas Teknologi Informasi Universitas Tarumanagara, yang dibuat oleh Ferryanto, mahasiswa Program Studi Teknik

Informatika Fakultas Teknologi Informasi Universitas Tarumanagara [2]. Aplikasi ini menggunakan metode *Bayesian Belief Networks* sebagai landasan teori dalam merancang aplikasi. Hasil penelitian menunjukkan bahwa aplikasi yang dibentuk mampu menampilkan rekomendasi rencana studi kepada mahasiswa dan mendeteksi mahasiswa yang terancam *drop out*.

2.3. Sistem Rekomendasi Perencanaan Studi Mahasiswa Dengan Menggunakan Algoritma Apriori Dan Naive Bayes

Penelitian terkait selanjutnya yang pernah disusun adalah “Sistem Rekomendasi Perencanaan Studi Mahasiswa Dengan Menggunakan Algoritma Apriori Dan *Naive Bayes*”, oleh Elizabeth Erlsha, mahasiswi Universitas Tarumanagara Fakultas Teknologi Informasi Jurusan Teknik Informatika [1]. Aplikasi ini menggunakan Algoritma Apriori untuk membentuk pola mata kuliah, sedangkan *Naive Bayes* diharapkan dapat memprediksi persentase kelulusan dalam rekomendasi perencanaan studi yang ditawarkan. Hasil yang ditawarkan oleh algoritma Apriori digunakan untuk merekomendasikan mata kuliah kepada mahasiswa dengan mencocokkan pola mata kuliah

3. METODE PENELITIAN

3.1 Pre-processing Dataset

Untuk mendapatkan hasil pengujian yang baik, pre-processing dataset dilakukan pada dataset pelatihan dan dataset pengujian. Dataset pelatihan yang digunakan adalah data nilai program studi Teknik Informatika angkatan 2008 sampai 2012 di salah satu perguruan tinggi swasta di Jakarta sebanyak 466 data. Sedangkan dataset pengujian yang digunakan adalah data nilai mahasiswa program studi Teknik Informatika angkatan 2013 dan 2014 di perguruan tinggi yang sama sebanyak 182 data.

Pre-processing dataset latih diawali dengan data cleaning yaitu membuang data mahasiswa yang belum lulus atau sudah lulus namun memiliki nilai mata kuliah yang tidak lengkap sesuai yang dibutuhkan. Proses *data cleaning* kedua dilanjutkan dengan menghapus nilai yang tidak diperlukan yaitu nilai mata kuliah yang tidak dibutuhkan dari data hasil data cleaning pertama. Setelah dilakukan *data cleaning*, didapatkan data latih siap pakai sebanyak 240 *instances*.

Mengingat kelemahan dari algoritma *VFIS* di mana jumlah data latih untuk setiap kelas harus seimbang, maka dilakukan *data reduction* untuk menyeimbangkan jumlah data untuk 3 kelas yang ada. Setelah dilakukan *data reduction* didapatkan data latih siap pakai sebanyak 135 *instances* yang terdiri dari 45 *instances* kelas 7 semester, 45 *instances* kelas 8 semester dan 45 *instances* kelas di atas 8 semester.

Pre-processing terhadap data uji dilakukan hanya dengan menyesuaikan tabel dengan struktur yang ada pada basis data.

3.2 Voting Feature Interval 5

Voting Feature Interval 5 adalah algoritma klasifikasi dengan konsep yang menjadi ciri khas utamanya yaitu adanya dari setiap fitur, di mana klasifikasi didasarkan pada nilai *vote* yang ada pada fitur itu.

Tahap – tahap pembelajaran pada algoritma ini dilakukan dengan,

1. Input data latih.
2. Pembentukan end point dari setiap kelas untuk setiap features.

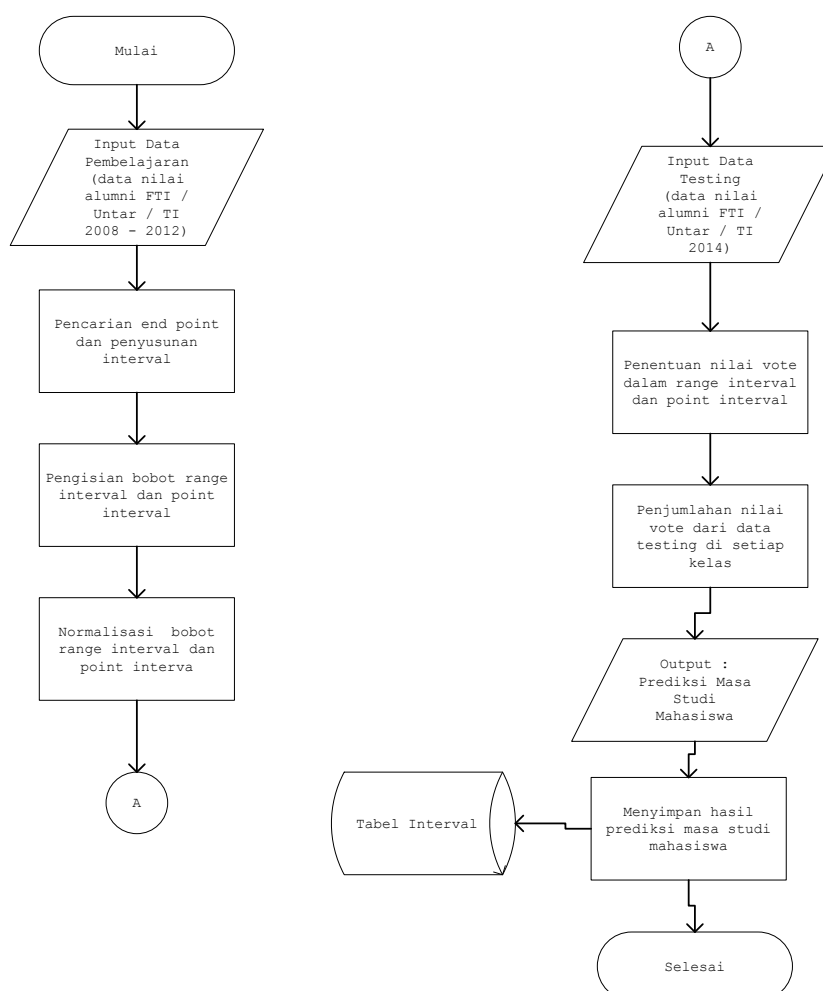
3. Pengisian frekuensi point interval dan range interval untuk setiap kelas pada setiap features.
4. Normalisasi frekuensi point interval dan range interval untuk setiap kelas pada setiap features, dengan rumus :

$$\frac{(\text{frekuensi} * \text{jumlah data kelas bersangkutan})}{\text{jumlah keseluruhan data}}$$

Sedangkan tahap – tahap pengujian pada algoritma ini dilakukan dengan,

1. Input data testing
2. Pencarian ruang interval yang tepat untuk masing – masing features, untuk menentukan frekuensi yang akan menjadi nilai vote pada features bersangkutan.
3. Perhitungan jumlah nilai vote total dari setiap interval yang terpilih.
4. Kelas dengan nilai vote terbesar menyatakan bahwa kelas tersebut adalah kelas dari data pengujian yang diinputkan.

Untuk penggambaran yang lebih jelas mengenai algoritma, dapat dilihat pada flowchart di gambar 1.



Gambar 1. Alur Aplikasi

3.3 K-Fold Cross Validation

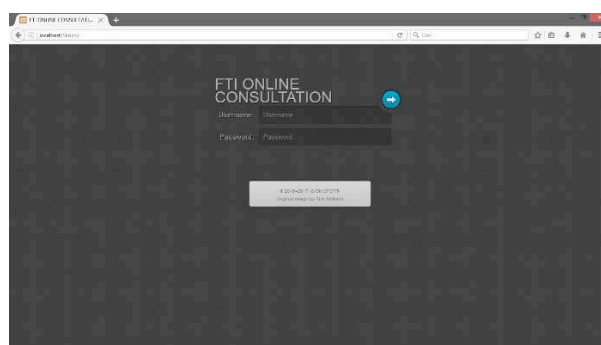
K-Fold Cross Validation adalah metode validasi dengan membagi data ke dalam *k-subset*, kemudian melakukan pengulangan sebanyak *k* kali untuk pembelajaran dan pengujian. Pada setiap pengulangan, digunakan satu *subset* sebagai data uji dan *subset* lainnya sebagai data pembelajaran. Keuntungan dari metode ini adalah setiap data, minimal akan menjadi data uji sebanyak satu kali dan akan menjadi data *learning* juga minimal sebanyak satu kali

4. ANALISIS DAN PEMBAHASAN

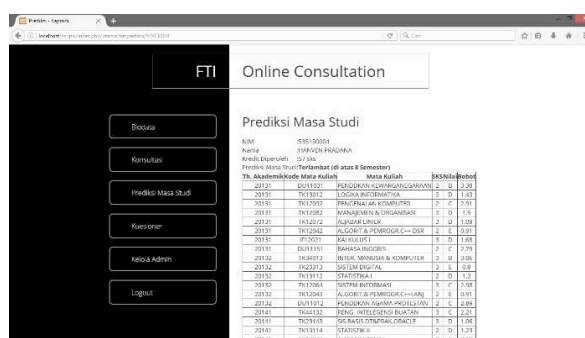
Pengujian yang dilakukan terhadap aplikasi konsultasi akademik online dengan fitur utama melakukan prediksi masa studi mahasiswa dilakukan melalui pengujian terhadap modul dan pengujian terhadap klasifikasi. Pengujian terhadap modul bertujuan menguji apakah setiap modul dalam aplikasi berjalan dengan baik. Pengujian terhadap klasifikasi, dilakukan untuk memastikan bahwa klasifikasi berjalan sesuai konsep.

Hasil pengujian terhadap modul menunjukkan bahwa semua modul berjalan dengan baik dan sesuai rancangan. Berikut ini merupakan modul dan submodul dalam aplikasi:

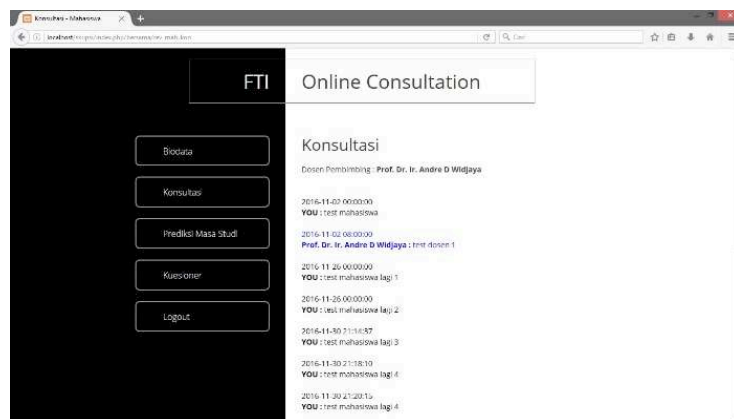
1. Modul *Login*
Hasil pengujian dapat dilihat pada gambar 2.
2. Modul *Prediksi*
Hasil pengujian dapat dilihat pada gambar 3.
3. Modul *Konsultasi*
Hasil pengujian dapat dilihat pada gambar 4.
4. Modul *Kuesioner*
Hasil pengujian dapat dilihat pada gambar 5.
5. Modul *Administrasi*
Hasil pengujian dapat dilihat pada gambar 6.



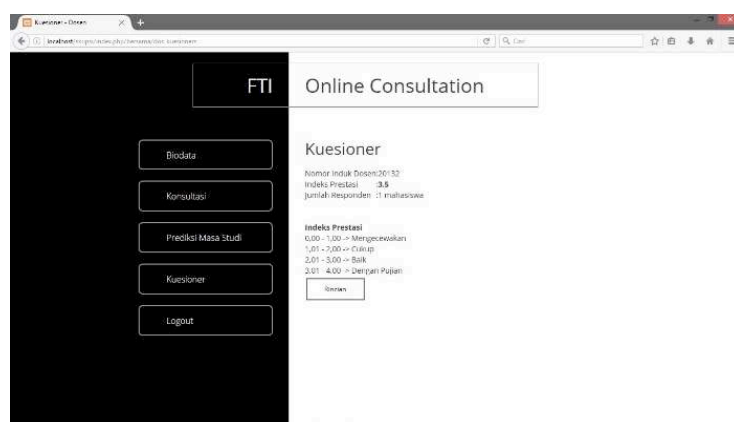
Gambar 2. Modul Login



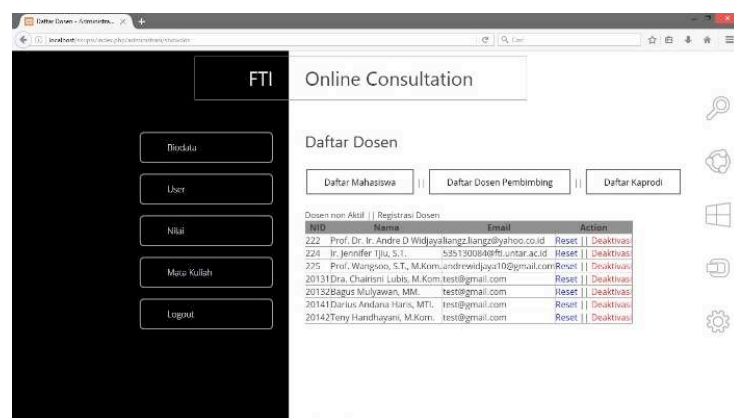
Gambar 3. Modul Prediksi



Gambar 4. Modul Konsultasi



Gambar 5. Modul Kuesioner



Gambar 6. Modul Administrasi

Setelah dilakukan pengujian terhadap modul dan didapatkan bahwa seluruh modul berjalan baik, selanjutnya pengujian dilakukan dengan pengujian terhadap klasifikasi. Percobaan terhadap klasifikasi dilakukan dengan 4 skenario percobaan sebagai berikut :

1. Percobaan I

Percobaan I dilakukan dengan menginputkan data dummy di mana data tersebut juga diklasifikasikan secara manual. Data ini terdiri dari 15 instances dengan 5 instances untuk masing – masing kelas, kemudian membandingkan hasil perhitungan manual dengan hasil komputasi dari aplikasi untuk memastikan bahwa perhitungan komputer berjalan benar. Hasil pengujian pada percobaan I yaitu aplikasi berhasil melakukan perhitungan end point, frekuensi range interval, frekuensi point interval dan nilai vote dengan tepat.

2. Percobaan II

Percobaan II dilakukan dengan menginput data latih utama yaitu 135 instances dengan 45 instances untuk masing – masing kelas ke dalam aplikasi di mana data tersebut sudah dibagi sesuai k-fold cross validation yang terbagi 5 subset dengan 27 instances tiap subsetnya. Tujuan dari pengujian ini adalah menghitung tingkat akurasi dari aplikasi yang dibangun dan didapatkan tingkat akurasi sebesar 73,33% dengan standar deviasi 4,829. Hasil pengujian terdapat pada tabel 1.

Subset Uji	Tepat	Meleset	Akurasi
A	19	8	70,37%
B	21	6	77,78%
C	21	6	77,78%
D	18	9	66,67%
E	20	7	74,07%

Tabel 1. Tabel Hasil Uji Percobaan II

3. Percobaan III

Percobaan III dilakukan dengan menginput data latih utama yaitu 60 instances ke dalam 4 subset yang berasal dari percobaan II namun sudah dibersihkan dari data outlier. Sebaran data yaitu 15 instances untuk setiap subset. Hasil pengujian dari percobaan III yaitu tingkat akurasi sebesar 93,33% dengan standar deviasi 5,44. Dari percobaan ini juga dapat disimpulkan bahwa metode Voting Feature Interval kurang robust terhadap data outlier. Hasil percobaan dapat dilihat pada tabel 2.

Subset Uji	Tepat	Meleset	Akurasi
W	14	1	93,33%
X	15	0	100%
Y	13	2	86,67%
Z	14	1	93,33%

Tabel 2. Tabel Hasil Uji Percobaan III

4. Percobaan IV

Percobaan IV adalah pengujian terhadap situasi nyata penggunaan aplikasi, yaitu dengan menginput 135 instances data latih dengan 45 instances untuk tiap kelasnya, kemudian menginputkan data uji yaitu sebanyak 91 instances untuk angkatan 2013 dan 91 instances untuk angkatan 2014. Dari 182 kali percobaan didapatkan 148 tepat dan 34 meleset..

5. KESIMPULAN

Kesimpulan yang dapat diperoleh dari hasil pengujian yang dilakukan terhadap aplikasi yang dibangun yaitu adalah sebagai berikut:

1. Seluruh modul yang ada di dalam aplikasi sudah berjalan dengan spesifikasi dan sudah berfungsi sebagaimana semestinya.

2. Aplikasi berhasil menghasilkan prediksi masa studi mahasiswa dengan tingkat akurasi yaitu 73,33% dengan standar deviasi 4,829 untuk percobaan dengan data nilai mahasiswa FTI Untar angkatan 2008 sampai 2012.
3. Aplikasi berhasil menghasilkan prediksi masa studi mahasiswa dengan tingkat akurasi yaitu 93,33% dengan standar deviasi 5,443 untuk percobaan dengan data nilai mahasiswa FTI Untar yang masa studinya sejalan dengan nilainya, angkatan 2008 sampai 2012.
4. Aplikasi berhasil menghasilkan prediksi masa studi mahasiswa secara tepat sebanyak 148 data uji dari percobaan dengan data uji mahasiswa FTI Untar program studi TI angkatan 2013 dan 2014 sebanyak 182 data uji.
5. Algoritma klasifikasi *Voting Feature Interval* 5 memiliki kekurangan dalam menangani data *outlier*.

Untuk pengembangan aplikasi lebih lanjut di waktu yang akan datang, hal – hal yang perlu diperhatikan adalah :

1. Melakukan klasifikasi berdasarkan nilai mata kuliah yang lebih dinamis, yaitu tidak hanya terbatas pada 25 mata kuliah.
2. Analisa kesalahan sistem yang dilakukan secara otomatis dan bukan analisis manual.
3. Penggunaan SMTP *Sendgrid* berbayar, agar email yang dikirimkan dapat dipastikan masuk ke *folder inbox*.
4. Melakukan klasifikasi dengan algoritma SVM untuk mendapatkan tingkat akurasi yang lebih tinggi, karena algoritma SVM lebih *robust* terhadap data *outlier*.

UCAPAN TERIMAKASIH

Penulis mengucapkan rasa terima kasih sebesar-besarnya kepada Tuhan Yesus Kristus karena hikmat dan rahmat yang dilimpahkan sehingga makalah ini dapat diselesaikan. Penulis juga tidak lupa mengucapkan terima kasih kepada **Ibu Lely Hiryanto, ST., M.Sc**, selaku dosen pembimbing I, **Ibu Teny Handhayani, S.Kom, M.Kom**, selaku dosen pembimbing II, seluruh dosen dan staff Fakultas Teknologi Informasi Universitas Tarumanagara, keluarga besar penulis, teman-teman penulis yaitu Vinesia Subyanca, Leonardo Cahaya, Hendrick Wijaya, dan Michael Ernest.

DAFTAR PUSTAKA

- [1] Erlsha, E. *Sistem Rekomendasi Perencanaan Studi Mahasiswa Dengan Menggunakan Algoritma Apriori Dan Naivebayes (Studi Kasus FTI Untar)*. Jakarta: Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara., 2014.
- [2] Ferryanto. *Aplikasi Perencanaan Studi Mahasiswa Menggunakan Bayesian Belief Networks (Studi Kasus: FTI Untar)*. Jakarta: Program Studi Teknik Informatika Fakultas Teknologi Informatika Universitas Tarumanagara (Skripsi tidak dipublikasikan)., 2014.
- [3] Demiroz, G. *Non-Incremental Classification Learning Algorithms Based on Voting Feature Intervals*. Diambil kembali dari <http://www.cs.bilkent.edu.tr/tech-reports/1997/BU-CEIS-9715.ps.gz>, 9 September 2016.
- [4] Guvenir, H. A. *A Classification Learning Algorithm Robust to Irrelevant Features*. Ankara: Department of Computer Engineering and Information Science., 1998.
- [5] Guvenir, H., Demiroz, G., & Ilter, N. Learning Differential Diagnosis of Erythematous-Squamous Diseases using Voting Feature Intervals. *Artificial Intelligence in Medicine*, Ankara: Artif Intell Med., 1998.

- [6] Mustafa, M. S. *Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining*. Makassar: Academia., 2012.
- [7] Schneider, J. (1997, February 7). *CS CMU*. Diambil kembali dari Cross Validation: <https://www.cs.cmu.edu/~schneide/tut5/node42.html>, 9 September 2016.
- [8] Valacich, George, J. S., F, J., & Hoffer, J. A. *Essentials of Systems Analysis and Design 4th Edition*. Upper Saddle River: Prentice Hall., 2009.
- [9] Witten, I. H., Frank, E., & Hall, M. H. *Data Mining: Practical Machine Learning Tools and Techniques 3rd Edition*. Burlington: Elsevier., 2011.
- [10] Liming Fu. *Neural Network in Computers Intelligence*. Singapura: McGraw Hill., 1994.